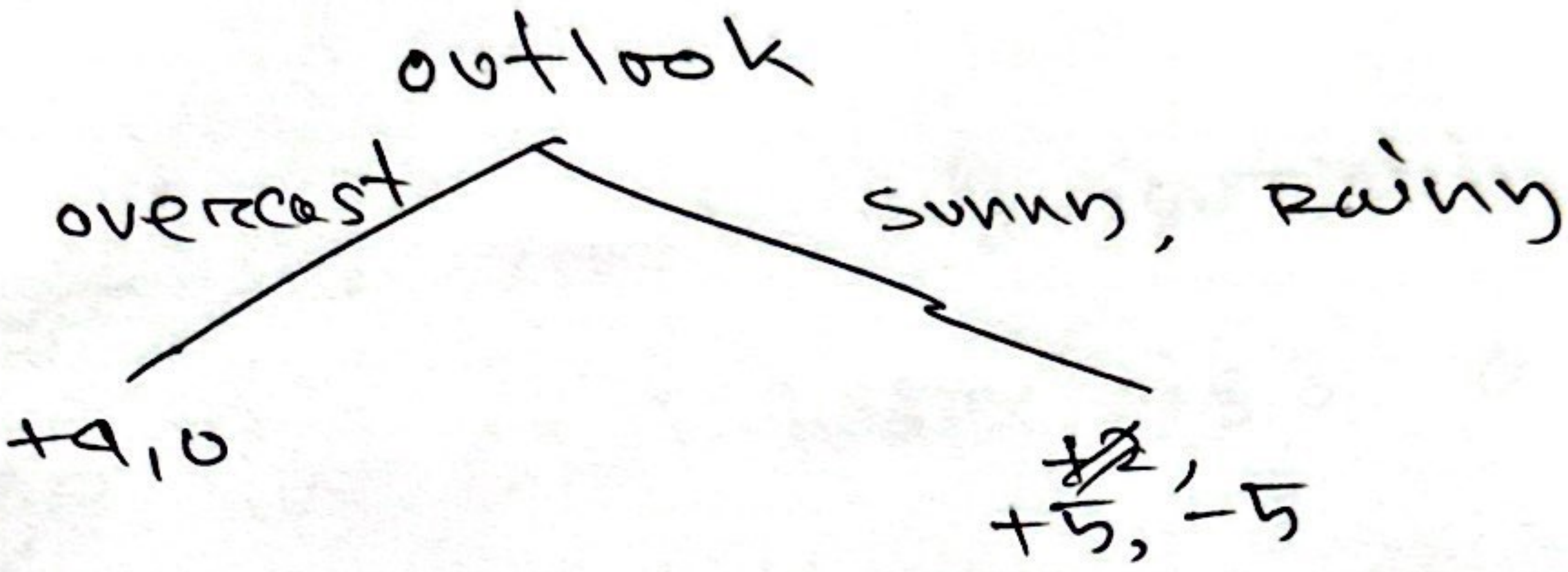
2) Gini index of outlook

Before using outlook

$$I_1 = 1 - \left(\frac{9}{14}\right)^2 - \left(\frac{5}{14}\right)^2$$

$$= 0.46$$

<u>After using outlook</u>

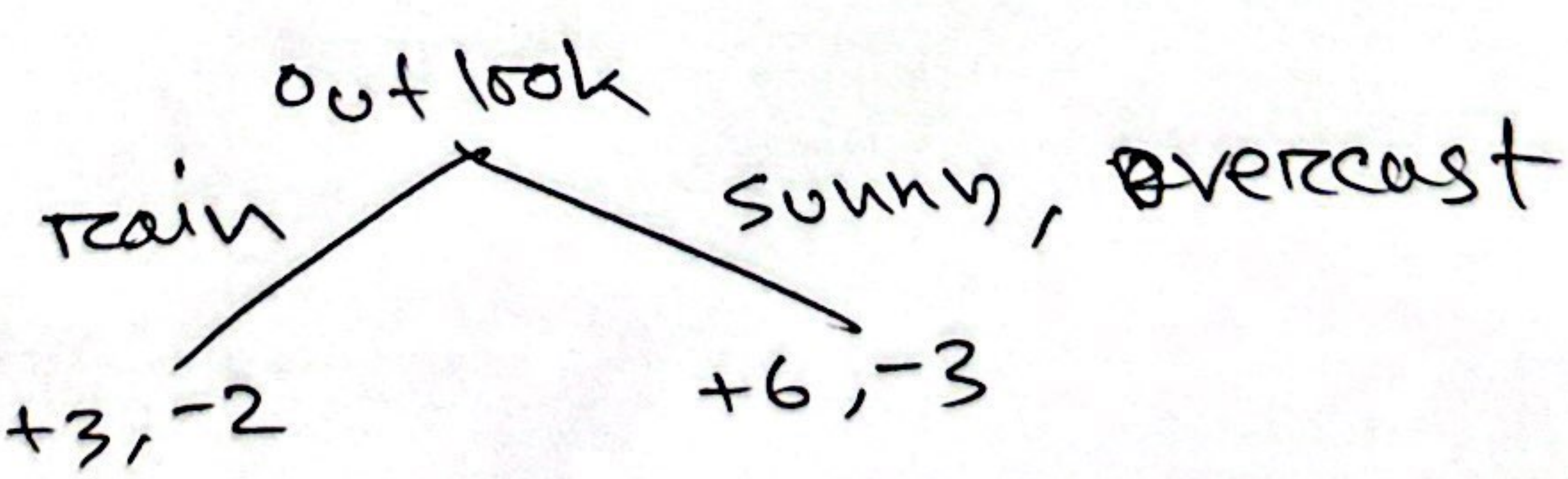


$$I_2 = \frac{4}{14}\left\{1-\left(\frac{4}{4}\right)^2-\left(\frac{0}{4}\right)^2\right\} + \frac{10}{14}\left\{1-\left(\frac{5}{10}\right)^2-\left(\frac{5}{10}\right)^2\right\}$$

$$= 0.36$$

$\therefore$ change in Gini index $= 0.46 - 0.36$

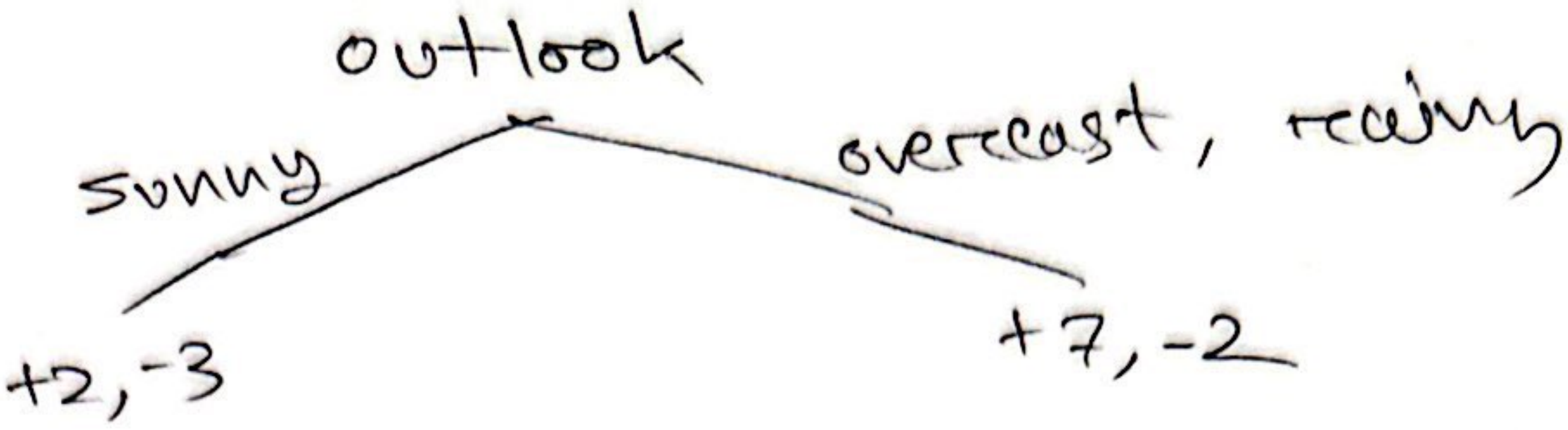
$= 0.10$



$$I_2 = \frac{5}{14}\left\{1-\left(\frac{3}{5}\right)^2-\left(\frac{2}{5}\right)^2\right\} + \frac{9}{14}\left\{1-\left(\frac{6}{9}\right)^2-\left(\frac{3}{9}\right)^2\right\}$$

$$= 0.46$$

Change in Gini index $= 0.46 - 0.46 = 0.$

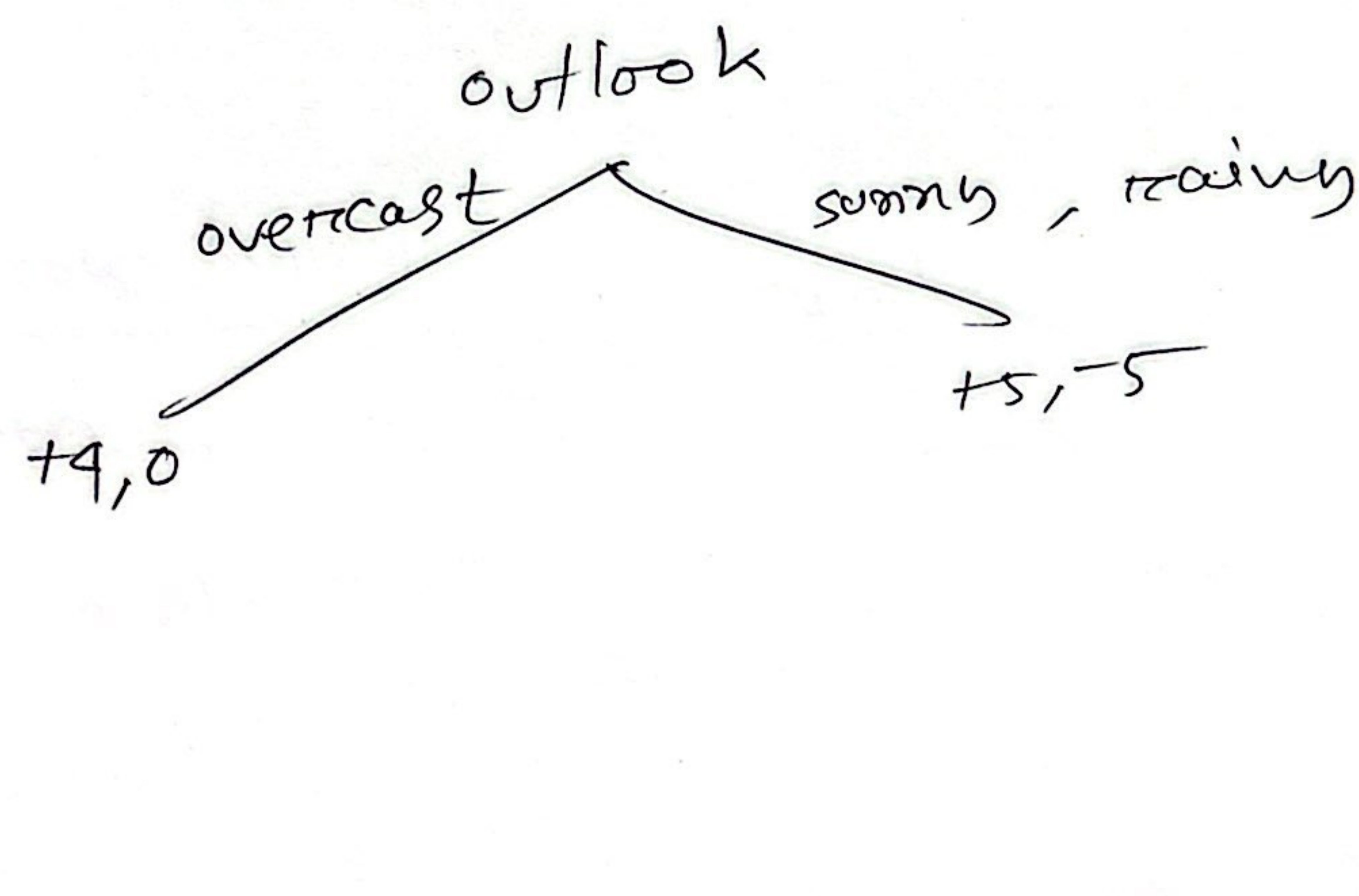


$$I_2 = \frac{5}{14}\left\{1-\left(\frac{3}{5}\right)^2-\left(\frac{2}{5}\right)^2\right\}+\frac{9}{14}\left\{1-\left(\frac{7}{9}\right)^2-\left(\frac{2}{9}\right)^2\right\}$$

$$= 0.39$$

Change in Gini Index $= 0.46 - 0.39$
$= 0.07$

so, for the root of outlook the change in Gini index is $= 0.36$



3) Rule — Outlook = sunny

Rule	Class	function
outlook = sunny	NO	$D1^{+}$, $D2^{-}$, $D8^{-}$, $D9^{+}$, $D11^{+}$

No = 3
Yes = 2
Total = 5

$$\text{accuracy} = \frac{3}{5} \times 100\%$$
$$= 60\%$$

$$\text{coverage} = \frac{5}{14} \times 100\%$$
$$= 35.71\%$$

FP (False positive) = 2 পাবে
FN (False Negative) = 0 পাবে

Rule:

i) outlook = sunny

~~+2~~,

Total	
+	= 9
-	= 5

after apply rules
(+2, -3) কিন্তু দেখছে class = NO
so, +3, -2

$$\text{Foil-Gain} = 3\left(\log_2 \frac{3}{3+2} - \log_2 \frac{9}{9+5}\right)$$
$$= -0.3$$

Scanned with CamScanner

+9, -5   +5, +9

outlook = sunny , Humidity = High   Class = No

+3, -2   +3, 0

F.G$_1$ = -0.3   F.G2 = ?

$$F.G2 = 3\left(\log_2 \frac{3}{3+0} - \log_2 \frac{3}{3+2}\right)$$

$= \cancel{3x}$ 2.211 Ans

4)

$$\overline{error}(M_1) = (30.5+32.2+20.7+20.6+31.0+41.0+27.7+26.0+21.5+26.0)/10$$

$$= 27.72$$

$$\overline{error}(M_2) = (22.4+14.5+22.4+19.6+20.7+20.4+22.2+19.4+16.2+35.0)/10$$

$$= 19.03$$

$$\left(\overline{error}(M_1) - \overline{error}(M_2)\right) = 27.72 - 19.03$$

$$= 8.69$$

$$error(M_1) - error(M_2) = 30.5 - 22.4 = 8.1$$

$$32.2 - 14.5 = 17.7$$

$$20.7 - 22.4 = -1.7$$

$$20.6 - 19.6 = 1$$

$$31.0 - 20.7 = 10.3$$

$$41.0 - 20.4 = 20.3$$

$$27.7 - 22.2 = 5.5$$

$$26.0 - 19.4 = 6.6$$

$$21.5 - 16.2 = 5.3$$

$$26.0 - 35.0 = -9$$

$$\mathrm{Var}(M_1 - M_2)$$

$$= \frac{1}{K} \sum \left[\left(\mathrm{err}(M_1)_i - \mathrm{err}(M_2)_i \right) - \left(\overline{\mathrm{err}}(M_1) - \overline{\mathrm{err}}(M_2) \right) \right]^2$$

$$= \frac{1}{K} \Big[(8.1 - 8.69)^2 + (17.7 - 8.69)^2 + (-1.7 - 8.69)^2 + (20.3 - 8.69)^2$$
$$+ (1 - 8.69)^2 + (10.3 - 8.69)^2$$
$$+ (5.5 - 8.69)^2 + (6.6 - 8.69)^2$$
$$+ (5.3 - 8.69)^2 + (-9 - 8.69)^2 \Big]$$

$$= \frac{724.973}{10}$$

$$= 72.5$$

$$\therefore t = \frac{\overline{\mathrm{err}}(M_1) - \overline{\mathrm{err}}(M_2)}{\sqrt{\frac{\mathrm{Var}(M_1 - M_2)}{K}}}$$

$$= \frac{8.69}{\sqrt{\frac{72.5}{10}}}$$

$$= 3.23$$

Scanned with CamScanner

for 9 degrees of freedom (10 round - 1)

and 5% significance level,

the critical value, $z = 1.26$

$\therefore \quad t = 3.23 > z = 1.26$

so, null hypothesis is rejected,

so, M1 and M2 has difference

and one is better.

As $\overline{err}(M_1) > \overline{err}(M_2)$

so M2 is better

6) OR function

Linear equation : $0.5x_1 + 0.5x_2 - 1.25 = 0$

Learning rate , $\alpha = 0.01$

hard threshold function

new weight = old weight + α (Target $-$ ~~prediction~~ $h(x)$) $\times$ input

As hard threshold

$\therefore$ $0.5x_1 + 0.5x_2 - 1.25 \geq 0$ then output = 1

$0.5x_1 + 0.5x_2 - 1.25 < 0$ then output = 0

OR

	x_1	x_2	Y
1)	0	0	0
2)	0	1	1
3)	1	0	1
4)	1	1	1

for 1st input

input = 0,0

output = 0 [target]

~~prediction~~ = $h(x) = 0$

$0.5 \times 0 + 0.5 \times 0 - 1.25$

$= -1.25 < 0$

so, $h(x) = 0$

No update

6) OR function

Linear equation : $0.5x_1 + 0.5x_2 - 1.25 = 0$

Learning rate, $\alpha = 0.01$

hard threshold function

new weight = old weight + α (Target − ~~prediction~~ $h(x)$) $\times$ input

As hard threshold

$\therefore$ $0.5x_1 + 0.5x_2 - 1.25 \geq 0$ then output = 1

$0.5x_1 + 0.5x_2 - 1.25 < 0$ then output = 0

OR

	x_1	x_2	Y
1)	0	0	0
2)	0	1	1
3)	1	0	1
4)	1	1	1

for 1st input

input = 0,0

output = 0 [target]

~~prediction~~ = $h(x) = 0$

$0.5 \times 0 + 0.5 \times 0 - 1.25$

$= -1.25 < 0$

So, $h(x) = 0$

No update

Scanned with CamScanner

for 2nd input

input = 0, 1

output = 1 [target]

$h(x) = 0$

$0 \times 0.5 + 1 \times 0.5 - 1.25$

$= -0.75 < 0$

$h(x) = 0$

$\therefore$ new weight$_1$ = $0.5 + 0.01(1-0) \times 0$

$= 0.5$

new weight$_2$ = $0.5 + 0.01(1-0) \times 1$

$= 0.51$

eq: $0.5x_1 + 0.51x_2 - 1.25 = 0$

for third input

input = 1, 0

output = 1

$h(x) = 0$

$0.5 \times 1 + 0.51 \times 0 - 1.25$

$= -0.75 < 0$

$h(x) = 0$

new weight$_1$ = $0.5 + 0.01(1-0) \times 1$

$= 0.51$

new weight$_2$ = $0.51 + 0.01(1-0) \times 0$

$= 0.51$

eq: $0.51x_1 + 0.51x_2 - 1.25 = 0$

For 4th input

input = 1, 1
output = 1
$h(x)$ = 0

~~0.51 × 0.5~~
$0.51 \times 1 + 0.51 \times 1 - 1.25$
$-0.23 < 0$
$h(x) = 0$

new weight$_1$ $= 0.51 + 0.01(1-0) \times 1$
$= 0.52$

new weight$_2$ $= 0.51 + 0.01(1-0) \times 1$
$= 0.52$

equation : $\boxed{0.52x_1 + 0.52x_2 - 1.25 = 0}$

Scanned with CamScanner